

Video Diffusion Models

Year: 2022 | Citations: 2124 | Authors: Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi

Abstract

Generating temporally coherent high fidelity video is an important milestone in generative modeling research. We make progress towards this milestone by proposing a diffusion model for video generation that shows very promising initial results. Our model is a natural extension of the standard image diffusion architecture, and it enables jointly training from image and video data, which we find to reduce the variance of minibatch gradients and speed up optimization. To generate long and higher resolution videos we introduce a new conditional sampling technique for spatial and temporal video extension that performs better than previously proposed methods. We present the first results on a large text-conditioned video generation task, as well as state-of-the-art results on established benchmarks for video prediction and unconditional video generation. Supplementary material is available at <https://video-diffusion.github.io/>